Employee Attrition Analysis

Murad Mammadov

Harvard Extension School

STAT E-100: Introduction to Quantitative Methods for the Social Sciences and Humanities

Professor Hidefusa Okabe

May 14, 2025

Abstract

This paper examines employee attrition using a structured dataset from IBM. The dataset contains 1,470 records with both categorical and numerical variables. Key categorical variables such as 'Attrition', 'Department', 'BusinessTravel', and 'JobRole' were analyzed along with continuous variables like 'MonthlyIncome' and 'YearsAtCompany'. The analysis aims to identify whether workplace and demographic factors are associated with employee turnover.

Descriptive analysis and inferential statistical tests were used to find a relationship between employee attrition and other categorical variables. A two-sample t-test was used to compare the average Monthly Income and Years at the Company between employees who left the company and those who stayed. Furthermore, a chi-square test of independence was conducted to examine the relationships between categorical variables such as Attrition and OverTime.

Preliminary findings suggest that employees who left tended to have lower income, shorter tenure, and were more likely to work overtime. These findings underscore the significance of early tenure support and equitable compensation in employee retention strategies.

Keywords: Employee attrition, IBM dataset, Monthly income, Years at company, Categorical variables, Business travel, Job role, Department, Overtime, Two-sample t-test, Chi-square test, Descriptive analysis, Inferential statistics, Employee turnover, Workforce retention, Compensation, Tenure support.

Introduction

As a business owner, I have always been interested in the factors that influence employee retention and turnover. One strategy that caught my attention was Zappos' "Pay-to-Quit" offer, which gives new employees a financial incentive to resign after training if they do not feel aligned with the company's values. This approach aims to filter out uncommitted individuals, reduce long-term costs, and strengthen organizational culture. The reasoning behind it resonated with me and prompted reflection on why some employees leave while others stay. Although employee attrition has not been a serious issue in my own company, I have remained curious about the underlying reasons behind the few departures that did occur. I began recording these cases and organizing the information into a table using categorical variables such as job role, department, and tenure. This hands-on experience inspired me to explore broader datasets on employee attrition in search of common patterns and influencing factors.

Replacing a separated employee may cost between 16% and over 213% of their annual salary, depending on the position (Boushey & Glenn, 2012). High employee turnover adversely affects employee morale and productivity, leading to significant financial costs and challenges in maintaining organizational competitiveness and success (SHRM, 2017; Dynamic Management, 2023; EPRA, 2022; Wharton School, 2019).

To remain competitive in the rapidly expanding global economy and to keep pace with technological advances requires a workforce with robust institutional knowledge; therefore, employee retention is of great importance to business and academic communities (Benko & Weisberg, 2007; Becker, 2007; The Future of Work 2020, 2007). Employee attrition is primarily attributed to two key factors: monthly income and job satisfaction. Employee salaries have been found to significantly influence turnover intention, as compensation plays a crucial role in employees' decisions to remain with or leave an organization (Fatima et al., 2023). It is therefore imperative for organizations to offer substantial and satisfactory compensation packages in order to retain competent employees, attract highly efficient talent, motivate staff toward greater diligence, and effectively achieve organizational objectives. Employee satisfaction with salary is widely recognized as a significant determinant of workplace behaviors, such as absenteeism and turnover intentions (Judge, Piccolo, Podsakoff, Shaw, & Rich, 2010).

Prior research has shown that job satisfaction is strongly and inversely associated with employees' intention to leave an organization (Egan, Yang & Bartlett, 2004; Lambert, Hogan & Barton, 2001; MacIntosh & Doherty, 2010; Schwepker, 2001; Silverthorne, 2004). This empirical evidence indicates that higher employee satisfaction reduces the likelihood of employees seeking employment with a different organization. Hence, categorizing employee satisfaction as a variable may be beneficial in determining the correlation between employee satisfaction and attrition.

This research utilizes a structured dataset comprising 1,470 employee records from IBM to examine the primary factors affecting employee attrition. The dataset encompasses both categorical and numerical variables, with a particular emphasis on *JobSatisfaction* and *MonthlyIncome* as potential predictors of turnover. Additional pertinent variables include *Attrition*, *Department*, *JobRole*, and *OverTime*. By analyzing the correlations between job satisfaction levels, income, and attrition status, the study aims to identify patterns that can guide effective employee retention strategies.

This paper examines two main research questions:

1. Is there a relationship between *JobSatisfaction* and employee attrition?
2. How does *MonthlyIncome* vary between employees who leave the company and those who remain?

This research uses descriptive statistics and inferential methods, such as chi-square tests and two-sample t-tests, to analyze job satisfaction and compensation's impact on employee retention. The results can inform HR professionals and organizational leaders aiming to address turnover through specific strategies.

Data and Methods

This project used the dataset and structure provided in a Kaggle notebook by Bachmann (n.d.), which included IBM HR employee records. This study utilizes the IBM Human Resources Analytics dataset, comprising **1,470 observations and 35 variables** that describe employees' demographic, professional, and organizational attributes. The dataset's absence of missing values renders it highly suitable for statistical analysis without requiring data imputation or cleaning. The variables are either categorical (factors) or numerical (integers), with no mixed or textual types.

The target variable is **Attrition**, a binary categorical variable indicating whether an employee has left the organization ("Yes") or remained ("No"). This variable is central to the study, with the primary objective being identifying factors significantly associated with employee attrition.

Key Variables of Interest

For this study, the analysis focuses primarily on:

- **JobSatisfaction** (ordinal categorical): A self-reported rating from 1 (low satisfaction) to 4 (high satisfaction).
- **MonthlyIncome** (continuous numerical): Employee's gross monthly earnings.

Other supporting variables include:

- **Age**
- **Gender**
- **YearsAtCompany**

Statistical Design and Methods

A quantitative analytical approach is adopted, utilizing descriptive and inferential techniques to explore relationships between attrition and key variables.

Descriptive Statistics

Frequency tables and summary measures describe the distribution of key attributes. Visual tools such as boxplots and histograms support interpretation.

Two-Sample Independent T-Test

A t-test was conducted to compare the **MonthlyIncome** of employees who left the company with those who stayed. This test determines if income levels are significantly associated with attrition status.

Chi-Square Test of Independence

To assess whether attrition is related to **JobSatisfaction** levels, a chi-square test is performed.

Descriptive statistics were computed using R Studio to determine the number and percentage of employees who left the organization. As shown in **Table 1**, the dataset consists of **1,470 employees**, of whom 237 (16%) left the organization and 1,232 (84%) remained. This indicates a class imbalance in the attrition variable, with most employees staying.

This imbalance should be considered in subsequent analyses, as it may affect the validity of statistical comparisons and the performance of predictive models. The 16% attrition rate suggests the significance of identifying factors, such as job satisfaction and compensation, that may contribute to turnover within a smaller portion of the workforce.

Table 1. Employee Attrition Summary

Total Employees	1470
Left (Yes)	237
Stayed (No)	1232
% Left	16%
% Stayed	84%

To assess whether there was a noticeable difference in Monthly Income between employees who left the organization and those who stayed, a two-sample t-test was utilized. This statistical method allows us to compare two independent variables. In other words, we will test whether to accept or reject the null hypothesis. The null hypothesis stated that the two groups had equal mean incomes ($\mu_1 = \mu_2$), while the alternative hypothesis proposed a difference in means ($\mu_1 \neq \mu_2$). Again, using R language, a t-test is accomplished as follows:

```
> t.test(MonthlyIncome ~ Attrition, data = data, var.equal = FALSE)

        Welch Two Sample t-test

data:  MonthlyIncome by Attrition
t = 7.4826, df = 412.74, p-value = 4.434e-13
alternative hypothesis: true difference in means between group No and group Yes is not equal to 0
95 percent confidence interval:
 1508.244 2583.050
sample estimates:
 mean in group No mean in group Yes
         6832.740          4787.093
```

As shown in the R output for two-sample t-test, there is a statistically significant difference in mean Monthly Income between the two groups, (412.74) = 7.48, $p < .001$. This means that employees who stayed in the organization had a higher average Monthly Income *(M=6832.74)* than those who left the organization *(M=4787.09).* The actual difference in means between these two independent groups is not equal to 0, which means the null hypothesis is rejected and the alternative hypothesis ($\mu_1 \neq \mu_2$) is accepted. The 95% confidence interval for the difference in means ranged from *1508.24 to 2583.05,* and this means that we are 95% confident that the actual difference in average Monthly Income between employees who stayed and left is between this range.

This finding suggests that income might be a significant factor in employee attrition, with lower-earning employees more likely to leave the company, given the very low p-value ($p < .001$), which is far below the standard threshold of .05 and which means the result is statistically significant, the likelihood of observing this difference by chance is extremely small, providing strong evidence against the null hypothesis.

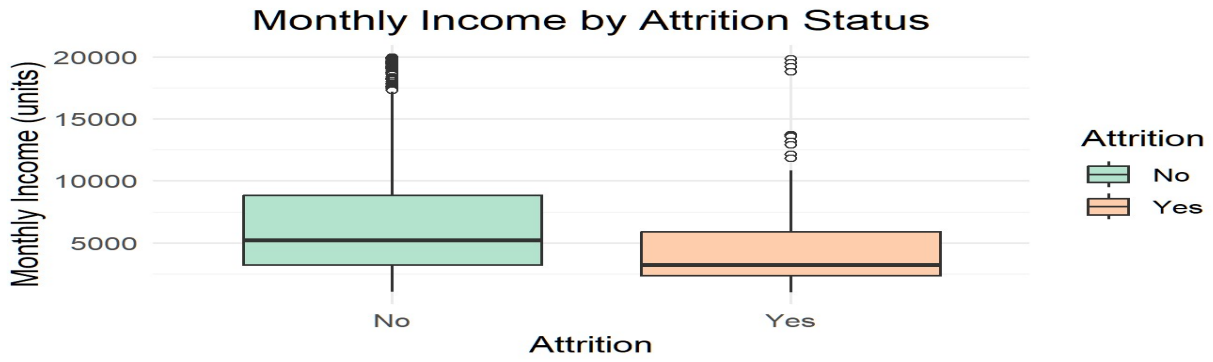
A boxplot was employed to visualize the distribution of Monthly Income between employees who left the organization and those who remained. As shown in Figure 1 and Table 2, the distribution of Monthly Income differed substantially between employees who left the organization and those who stayed. Employees who stayed exhibited a wider interquartile range (IQR = 5,623), indicating a broad spread of incomes within the middle 50% of this group. On the other hand, those who left had a narrower income distribution (IQR = 3,543). The lower median and compressed IQR among those left suggest more uniform and lower earnings.

Furthermore, the "No" group displayed more high-income outliers, further contributing to the group's higher income variability. These boxplot patterns visually reinforce the independent samples t-test findings, which identified a statistically significant difference in average Monthly Income between the two attrition groups, $t(412.74) = 7.48$, $p < .001$.

Table 2. Interquartile Ranges (IQR) of Monthly Income by Attrition Status

Values	
data	num [1:10] 1 2 3 NA 5 6 7 8 9 10
iqr_no	5623
iqr_value	5
iqr_yes	3543

Figure 1. Distribution of Monthly Income by Employee Attrition Status



Other supporting categorical variables, such as Age, Gender, and YearsAtCompany, were also tested independently to examine their relationship with attrition. As shown in **Figure 3**, the boxplot indicates that employees who left the organization tended to be younger on average, with a lower median age and narrower interquartile range compared to those who stayed. This suggests that younger employees, particularly early-career professionals, may be more likely to leave, which should be considered while making potential hirings. As further illustrated in **Figure 5**, employees who left also had shorter tenures at the company, suggesting that attrition is more likely among newer hires. This finding may align with the phenomenon that the longer employees work at a company, the more loyal they become and the less likely they are to leave the organization.

Additionally, this finding underscores the importance of early engagement and retention strategies. Meanwhile, **Figure 4** shows that attrition by gender is relatively balanced, though slightly higher among males in absolute numbers; however, no noticeable gender-based disparity is observed visually.

To sum up, these patterns support the broader generational inference that **millennials**, who are generally younger and have shorter tenure, are more likely to seek alternative employment

opportunities early in their careers. Additionally, **older employees (boomers)** may show elevated attrition as they approach retirement, explaining their position as the second-highest turnover group. These visual trends indicate that early-career mobility and late-career exit affect attrition rates across ages.

Figure 3. Boxplot of Age Distribution by Attrition Status

Figure 4. Bar Chart of Attrition by Gender

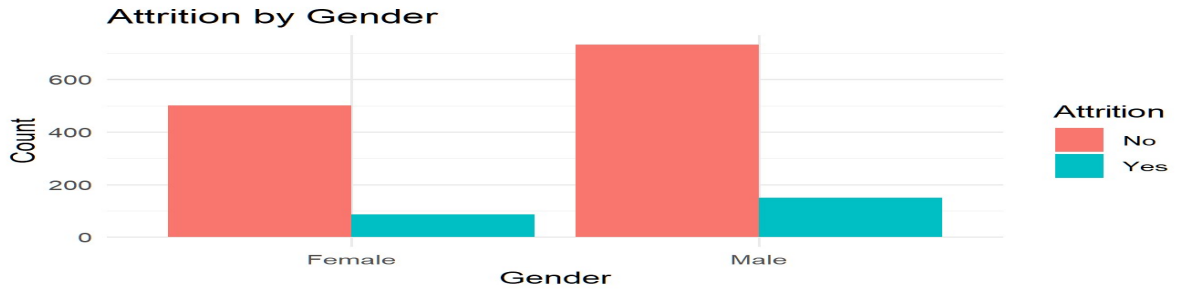
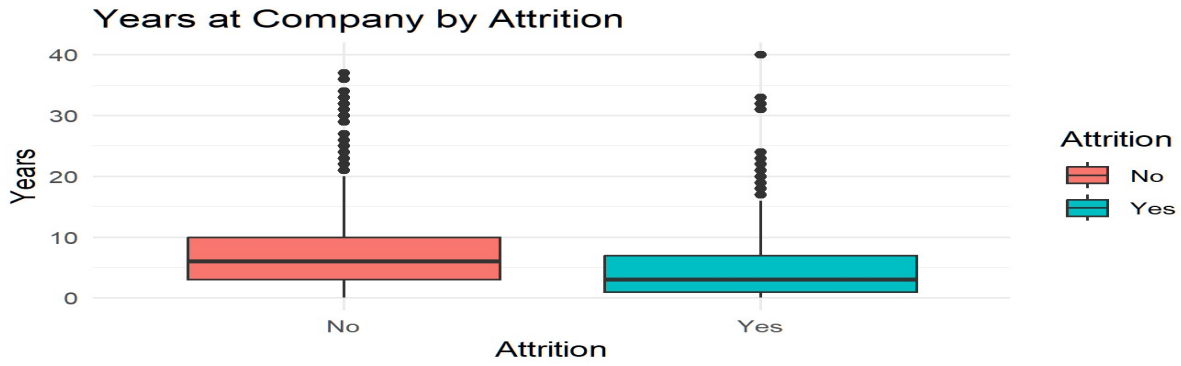


Figure 5. Boxplot of Years at Company by Attrition Status

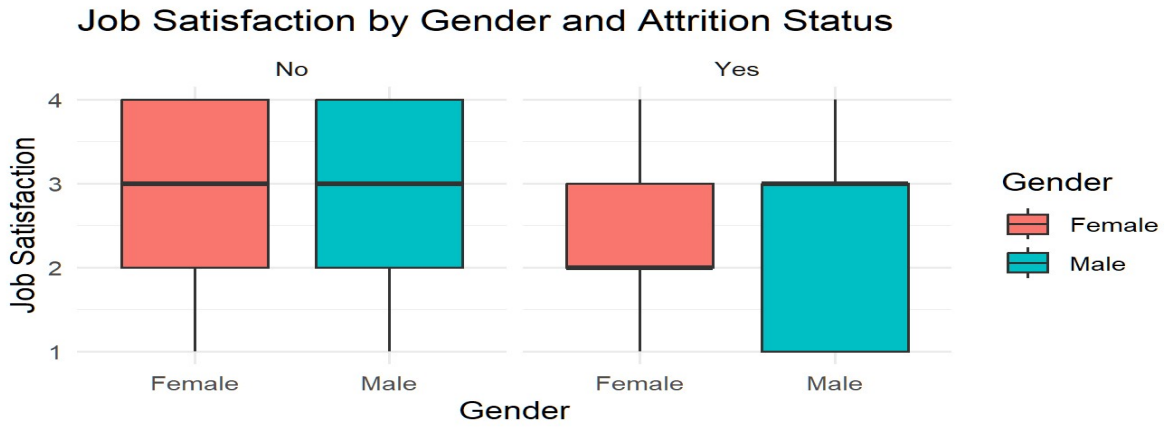


While low job satisfaction is widely recognized as contributing to employee attrition, it does not fully account for turnover behavior. Some employees with low satisfaction levels may still choose to remain with the organization, while others may leave despite reporting a moderate or high level of satisfaction. This complexity suggests that job satisfaction interacts with other individual-level characteristics, such as age, gender, and years at the company, in shaping

turnover outcomes. Therefore, it is essential to examine how job satisfaction varies across these factors and how it correlates with attrition.

The relationship between Job Satisfaction and Attrition was examined using Chi-square analysis and one-way Analysis of Variance (ANOVA) to investigate further the factors associated with employee attrition. The chi-square test assessed whether attrition status was independent of job satisfaction levels, treating both as categorical variables. On the other hand, the ANOVA tested whether the mean job satisfaction scores differed significantly between employees who left the organization and those who remained. To examine the relationship between job satisfaction and attrition, it might be better to start by examining whether one gender is more dissatisfied. As shown in Figure 6, job satisfaction levels are virtually unchanged for individuals who remained with the organization. However, females reported lower satisfaction levels among those who left the organization than males. This finding might indicate that, though slightly, dissatisfied female employees are more likely to leave the organization than males.

Figure 6. Job Satisfaction by Gender and Attrition Status



To further support, or examine, the previously mentioned assumption (This finding might serve as a predictor that, though slightly, dissatisfied female employees are more likely to leave the organization in comparison with males), one-way ANOVA statistical test was employed to see if there is a statistically significant difference in average job satisfaction between male and female employees. The results indicated that the difference in mean job satisfaction between genders was not statistically significant, $F(1, 1468) = 1.63$, $p = .203$. This suggests that gender is not a significant predictor of job satisfaction in this sample.

Table 3. One-Way ANOVA: Job Satisfaction by Gender

```
> anova_gender <- aov(JobSatisfaction ~ Gender, data = data)
> summary(anova_gender)
              Df Sum Sq Mean Sq F value Pr(>F)
Gender         1      2   1.976   1.625  0.203
Residuals   1468   1785   1.216
>
```

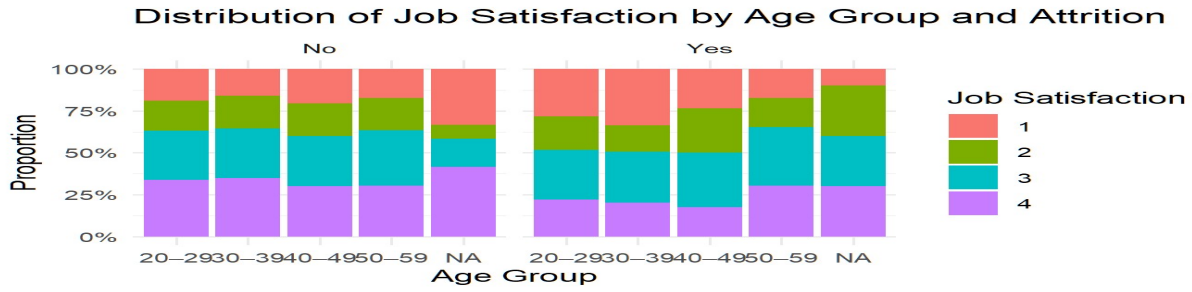
Other categorical variables, such as Age and MonthlyIncome, were analyzed to observe their interaction with job satisfaction. A Contingency Table of Counts was utilized to identify the least satisfied age group, which will be compared to see if this age group corresponds to the highest attrition group. As shown in Table 4, the 20–29 age group reported the highest percentage of very dissatisfied employees (21.51%), indicating that younger employees may be more vulnerable to dissatisfaction. This finding will explore whether this group demonstrates a higher likelihood of turnover, thus reinforcing the link between dissatisfaction and attrition.

Table 4. Dissatisfied Employees by Age Group

Age group	% Level 1 (Very Dissatisfied)
20–29	**21.51%**
30–39	18.26%
40–49	20.81%
50–59	18.88%

A stacked bar chart was generated using R Studio to examine whether the most dissatisfied group matches with the highest attrition rate. As shown in Figure 7, the 20-29 age group had the highest proportion of employees who left the organization. As shown in Table 4, this group also shows a high proportion of employees reporting lower job satisfaction (Levels 1 and 2), suggesting that younger employees may be more likely to leave due to dissatisfaction than other age groups.

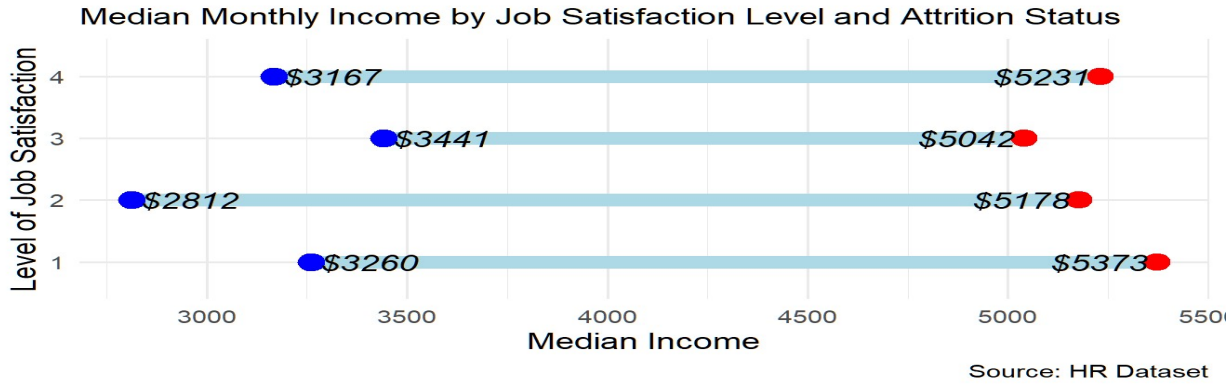
Figure 7. Job Satisfaction by Age Group and Attrition



Another valuable insight might be to analyze whether there are significant changes in the income level by JobSatisfaction. That is, are the individuals with lower satisfaction getting much less salary than the ones who are more satisfied? This can be valuable to identify how these variables, namely JobSatisfaction, Attrition, and MonthlyIncome, interact, especially to see if job satisfaction is sufficient to retain employees. A dumbbell plot was employed to visualize this correlation. As illustrated in Figure 8, across all satisfaction levels (1-4), employees who left the organization consistently have lower incomes than those who stayed. The most significant gap appeared at Level 2 ($2366), which suggests that low income may increase dissatisfaction,

contributing to higher attrition. Surprisingly, some employees with high satisfaction levels (3 and 4) still leave the organization, especially if their income is lower than that of satisfied employees who stay. This indicates that job satisfaction alone is insufficient to retain employees without competitive compensation.

Figure 8. Median Monthly Income by Job Satisfaction Level and Attrition Status



Finally, a Chi-square test of independence was used to assess the relationship between job satisfaction and attrition. The result was statistically significant, $X^2(3, N = 1470) = 17.51$, $p < .001$, indicating a meaningful association between job satisfaction and attrition status. Employees with lower job satisfaction were likelier to leave the organization than those with higher satisfaction levels. Again, since the p-value=.00056 is less than .05, the likelihood of observing a causal observation is very low.

```
> table_gender_js <- table(data$Attrition, data$JobSatisfaction)
> chisq.test(table_gender_js)

        Pearson's Chi-squared test

data:  table_gender_js
X-squared = 17.505, df = 3, p-value = 0.0005563

> print(table_attrition_js)

       1   2   3   4
  No  223 234 369 407
  Yes  66  46  73  52
> |
```

Discussion and Conclusion

Using the IBM Human Resources dataset, this study investigated the relationship between employee attrition and key organizational variables. The findings strongly suggest that Monthly Income and Job Satisfaction are significant factors associated with employee turnover. Overall, employees who left the organization had significantly lower monthly incomes and were more likely to report low satisfaction. Different statistical analyses validated that these two variables, although commonly mentioned, affect employee turnover.

Interestingly, even some employees with higher satisfaction scores chose to leave the company, especially when their salaries were lower than those of peers who stayed. This reinforces the idea that job satisfaction alone cannot retain talent without competitive compensation. The dumbbell plot provided visual support for this pattern.

Demographic trends were also carefully analyzed: younger employees (20–29 years) reported higher dissatisfaction and attrition rates, likely due to early-career mobility. Gender differences were minimal, though females who left reported slightly lower satisfaction. These findings highlight the need for targeted retention strategies, such as equitable compensation and early-career support programs.

While the analysis yielded valuable insights, some limitations must be considered. First, the dataset does not include qualitative variables such as reasons for leaving, job performance, work-life balance, or employee engagement—all of which may influence turnover. Second, the dataset represents employees from a specific industry sector. As such, the findings may not be generalizable to organizations in different sectors with varying work cultures, compensation norms, or employee engagement practices.

Future studies focusing on samples from multiple industries and organization types and covering other key independent variables could help better understand the reasons behind employee turnover. This will allow us to understand how employee turnover differs depending on work context, industry, and other industry-related circumstances. Furthermore, longitudinal data may be helpful in deeply exploring causality because tracking employee attitudes and behaviors across multiple time points would allow for stronger inferences about causality and attrition patterns.

References

1. Bachmann, J. (n.d.). Attrition in an organization - why workers quit [Kaggle notebook]. Kaggle. https://www.kaggle.com/code/janiobachmann/attrition-in-an-organization-why-workers-quit
2. Boushey, H., & Glenn, S. J. (2012, November). *There are significant costs to replacing employees*. Center for American Progress. https://www.americanprogress.org/wp-content/uploads/2012/11/CostofTurnover.pdf
3. Society for Human Resource Management (SHRM). (2017). *Retaining talent: A guide to analyzing and managing employee turnover*. Retrieved from https://www.shrm.org/content/dam/en/shrm/topics-tools/news/Retaining-Talent.pdf
4. Benko, C. &Weisberg, A. (2007). Implementing a Corporate Career Lattice: The Mass Career Customization Model." *Strategy & Leadership, 35*(5), 29-36.
5. Becker, F. (2007). Organizational Ecology and Knowledge Networks. *California Management Review, 49*(2), 42–61.
6. Fatima, S., Hina, M., & Lodhi, S. A. (2023). *Employee salary and employee turnover intention: A key evaluation considering job satisfaction and job performance as mediators*. *International Journal of Management Research and Emerging Sciences*. Retrieved from https://www.academia.edu/98663305
7. Judge, T. A., Piccolo, R. F., Podsakoff, N. P., Shaw, J. C., & Rich, B. L. (2010). *The relationship between pay and job satisfaction: A meta-analysis of the literature*. *Journal of Vocational Behavior*, *77*(2), 157–167. https://doi.org/10.1016/j.jvb.2010.04.002
8. Egan, T. M., Yang, B., & Bartlett, K. R. (2004). Organizational learning culture and job satisfaction affect motivation to transfer learning and turnover intention. *Human Resource Development Quarterly*, *15*(3), 279–301. doi:10.1002/hrdq.1104
9. Lambert, E. G., Hogan, N.L., & Barton, S. M. (2001). The impact of job satisfaction on turnover intent: a test of a structural measurement model using a national sample of workers. *The Social Science Journal*, *38*(2), 233–250. doi:10.1016/S03623319(01)001100
10. MacIntosh, E. W., & Doherty, A. (2010). The influence of organizational culture on job satisfaction and intention to leave. *Sport Management Review*, *13*(2), 106–117. doi:10.1016/j.smr.2009.04.006
11. Schwepker, C. H. (2001). Ethical climate's relationship to job satisfaction, organizational commitment, and turnover intention in the salesforce, *Journal of Business Research*, *54*(1), 39–52. doi: 10.1016/j.bbr.2011.03.031
12. Silverthorne, C. (2004). The impact of organizational culture and person-organization fit on organizational commitment and job satisfaction in Taiwan. *Leadership & Organization Development Journal*, *25*(7), 592–599. doi:10.1108/01437730410561477
13. OpenAI. (2023). *ChatGPT* (May 3 version) [Large language model]. https://chat.openai.com